STN PUB
NE

GILBERT

STATION'S QUARTERLY REPORT

1st Quarter

Calendar Year 1955

Sample size in stratified sampling

Northeastern
Forest Experiment Station
Upper Darby, Pa.

For In-Service Use Only

TABLE OF CONTENTS

	Page
Biometry	1
Forest Disease Research	4
Forest Economics	6
Forest Insect Research	8
Forest Utilization Service	10
Adirondack Research Center	18
Allegheny Plateau Research Center	22
Coastal Oak Pine Research Center	24
Kingston Research Center	25
Mountain State Research Center	29
Penobscot Research Center	33
Virginia Pine-Hardwood Research Center	37
White Pine-Hardwood Research Center	45
Hopkins Memorial Experimental Forest	50
Massabesic Experimental Forest	53

BIOMETRY

by C. A. Bickford

Sample size in stratified sampling (neglecting consideration of cost)

The problem of determining proper sample size for estimating distribution of volume by classes of quality, or an average quality index, was posed by the White Pine-Hardwoods Center. This led to the development of a procedure that might be used in other situations and is described below.

Data are available representing about 1000 trees showing for each the species, dbh, and merchantable height in standard 16 ft. logs; in addition, the class or grade of each log was recorded. Thus it was possible to obtain a quality index for each tree. (This was, in effect, an average price per unit of volume.)

The first question was whether or not stratified sampling would materially reduce the size of sample required. The answer is obtained by using the analysis of variance and computing a variance ratio. For convenience in presentation, we will first consider the case where there is only one basis of classification, say dbh. The usual procedure with unequal numbers of observations leads to estimates of mean square within and between diameter strata and for the total of all data. The dependent variable in this case was the quality index on a tree basis. The ratio usually computed to detect significant differences is mean square between classes divided by mean square within classes. Here, however, to determine if there is advantage to stratified sampling, divide total mean square by mean square within classes. If this ratio is greater than one and statistically significant, the case for stratified sampling is strong. In the absence of significance of this ratio, if the "between" over "within" is significant, the "total" over "within" is relatively large, and a sample is relatively expensive, there may still be an advantage to stratified sampling.

Once the variance ratios are known, the next step is to calculate sample size. In a homogeneous population where all variances are equal, e.g. no variance ratio significant, the usual formula applies:

$$n = \frac{s^2t^2}{e^2} \quad \text{if the population is very large}$$

$$\text{or} \qquad n = \frac{Ns^2t^2}{Ne^2 + s^2t^2} \quad \text{if the population is limited.}$$

Where n is sample size, N is number of possible samples, s is estimated standard deviation, e is maximum allowable error, and t is a confidence factor usually taken as 2.

When stratified sampling is appropriate the procedure is more complex. It is necessary first to establish a basis of weighting to distribute numbers of samples by classes. The simplest procedure is to use a uniform sampling fraction (taking one in k as a sample). This amounts to weighting on the basis of numbers of trees by classes and may not be much better than simple random sampling.

Another alternative is to weight in proportion to class means. It is also possible to weight in proportion to class standard deviations. In either case, the first step is to obtain approximate sampling fractions in the desired proportions to one another; the next is to calculate a correction factor that will result in the proper number of samples by classes. The procedure is briefly sketched below for the problem of determining quality index.

Asquality index is average price per unit volume, it is necessary to weight tree quality index, and average quality index by class, by volume. Thus proportions were calculated for class volume over total volume; these proportions are symbolized below as "h" (in other situations, "h" might be defined on the basis of number, or value, etc.).

Next obtain as good an estimate as is convenient of the standard deviation for each class; call this "s". It may be computed from a preliminary sample, estimated from the range, or guessed from similar related experience.

Now the maximum possible size of a sampling fraction is unity--you cannot sample more trees than there are in the population. Assume the sampling fraction of the class with the largest estimated "s" is one and calculate approximate sampling fractions for the other classes. If preferred, the method would also work if we start with an assumed sampling fraction of 1/2 instead of one. Call these fractions "F". Now if we use "e" to represent maximum allowable error (in basic units), the correction factor, c, is estimated from:

$$c = \frac{\text{sum of } \left(\frac{s^2h}{F}\right)}{e^2 + \text{sum of } s^2h}$$

Desired number of samples for each class may then be estimated from NcF, where N is total number of individuals in the class. The adjusted sampling fraction, "f", is cF. Strict observance of these numbers and fractions may be awkward or cumbersome. Fortunately it is not necessary as rounding off to more convenient numbers or fractions, if not overdone, affects the sampling error but very little.

If it is more convenient to express sampling error and allowable error as a percentage, this may be done, with appropriate modification of the formula for the correction factor.

Calculations following this procedure are not available but should be reported later by the Center.

FOREST DISEASE RESEARCH

A revised and improved staining technique for the identification of the white pine blister rust fungus in the bark and wood of infected trees has been prepared by Waterman. This report will be published in a technical journal or as a Station Research Note.

A manuscript by Waterman, entitled "The Relation of Valsa kunzei to cankers on conifers," has been submitted for publication in PHYTOPATHOLOGY. This paper describes three varieties of Valsa kunzei and gives the host range and parasitism of each. Control measures are discussed.

Phomopsis canker of Douglas fir, a serious disease on this tree in European plantations, was the reason for the establishment several years ago of a U.S.D.A. quarantine against certain conifers grown in Europe (Nursery Stock, Plant and Seed Quarantine No. 37). Hahn's research on this disease has disclosed that it is present but of relatively minor importance on several conifers in eastern United States. A statement justifying the lifting of this quarantine was prepared by Hahn and submitted to the Plant Quarantine Branch, Agricultural Research Service.

Hahn has completed a manuscript on the role of fungi as causal agents of birch dieback. None of the fungi isolated from the crowns, trunks, and roots of deteriorated birch in Maine and New Hampshire can be considered as primary causes of the disease. This paper will be submitted for publication as a Station Paper or in a technical journal.

The results of a study on incipient decay in Sitka spruce, made by Waterman and Hansbrough during and following World War II, have been revised for publication as a Station Paper. The data were previously available only in an Office Report, unpublished, but requests for detailed information on the techniques developed and the results obtained now seem to justify publication. This study includes the classification of the effect of seven species of fungi on the strength properties of wood from Sitka spruce logs, the development of a decay rating based upon the microscopical condition of the wood, and a discussion of the value of the microscopical rating in assessing strength losses.

Final arrangements have been made for a cooperative study with the Green Mountain National Forest of the relation of heart rot in red spruce to site, stand, and tree factors. Field work will start in May. Anyone interested in the details of this study should write to Mook at the Forest Insect and Disease Laboratory.